

# GONUTS: the Gene Ontology Normal Usage Tracking System

Daniel P. Renfro<sup>1</sup>, Brenley K. McIntosh<sup>1</sup>, Anand Venkatraman<sup>1</sup>, Deborah A. Siegele<sup>2</sup> and James C. Hu<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, Texas A&M University and Texas AgriLife Research and

<sup>2</sup>Department of Biology, Texas A&M University, College Station, TX 77843-3258, USA

Received August 13, 2011; Revised September 29, 2011; Accepted October 6, 2011

## ABSTRACT

The Gene Ontology Normal Usage Tracking System (GONUTS) is a community-based browser and usage guide for Gene Ontology (GO) terms and a community system for general GO annotation of proteins. GONUTS uses wiki technology to allow registered users to share and edit notes on the use of each term in GO, and to contribute annotations for specific genes of interest. By providing a site for generation of third-party documentation at the granularity of individual terms, GONUTS complements the official documentation of the Gene Ontology Consortium. To provide examples for community users, GONUTS displays the complete GO annotations from seven model organisms: *Saccharomyces cerevisiae*, *Dictyostelium discoideum*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Mus musculus* and *Arabidopsis thaliana*. To support community annotation, GONUTS allows automated creation of gene pages for gene products in UniProt. GONUTS will improve the consistency of annotation efforts across genome projects, and should be useful in training new annotators and consumers in the production of GO annotations and the use of GO terms. GONUTS can be accessed at <http://gowiki.tamu.edu>. The source code for generating the content of GONUTS is available upon request.

## GENE ONTOLOGY AND THE NEED FOR A GRANULAR USAGE GUIDE

The Gene Ontology has become a standard for the consistent functional annotation of genes and gene products

across all organisms (1–3). Well-established model organism databases (MODs) devote considerable resources to providing high-quality GO annotation based on manual curation from the experimental literature. Annotation of genes from other organisms is often inferred on the basis of homology to one of the established model organisms.

Two situations inspired us to build a general system to support community GO annotation. First, several well-established bacterial systems with extensive experimental literature (e.g. *Escherichia coli* and *Bacillus subtilis*) were not well covered by GO annotations based on direct experimental evidence. Second, the wealth of new complete genome sequences includes many organisms where there is significant experimental literature, but the research communities are too small to support a model organism database.

In some of these cases, the expertise in the experimental corpus is held by a small number of researchers focusing on specific biological problems. These experts could provide the most scientifically accurate and rich GO annotation and could provide useful input about the structure of the ontology itself, provided they familiarize themselves with the Gene Ontology. Typically these experts are only interested in a small number of genes and have little incentive to familiarize themselves with the intricacies of GO. Alternatively, researchers that are intimately familiar with GO have the potential to create many useful annotations, but typically do not have the necessary expertise to optimally annotate the wide variety of subject areas in the scientific literature. Either professional curators need to learn many new areas of biology, or experts need to become adequately acquainted with GO to provide quality annotations via community curation.

Training of large numbers of community members in GO annotation would be aided by more detailed documentation of best practices and pitfalls of GO. The Gene Ontology is structured as a directed acyclic graph

\*To whom correspondence should be addressed. Tel: +1 979 862 4054; Fax: +1 979 845 9274; Email: [jimhu@tamu.edu](mailto:jimhu@tamu.edu)  
Present address:

Anand Venkatraman, Monsanto Company, St Louis, MO 63167, USA.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

(DAG) where documentation can be applied to an entire branch, or more optimally, at the level of individual terms. Building such detailed documentation is a Herculean task for any small group, such as the GO consortium. However, we reasoned that large-scale documentation could be incrementally built through community collaboration, especially if community members record the nuances they encounter while learning the annotation process. The nature of wiki-based systems allows revisions and corrections of false starts and errors in the learning process. A wiki provides an exceptional environment for this type of self-documenting system, as well as valuable resources such as revision history and a dedicated page for discussion.

To capture detailed usage notes dealing with GO terms and to aid the annotation of specific genes by research experts, we constructed the Gene Ontology Normal Usage Tracking System (GONUTS), a wiki-based GO browser and community annotation system (<http://gowiki.tamu.edu>). GONUTS currently supports two main kinds of wiki-based content: (i) pages to capture community usage notes for specific GO terms and (ii) editable pages displaying editable GO annotations for specific genes.

## GONUTS AS A GENE ONTOLOGY TERM BROWSER

**Figure 1** shows an example of a GONUTS page for a GO term. GO terms are represented by MediaWiki Category pages, which like biological ontologies, can be represented as DAGs. Each page contains information about the GO term from the publicly downloadable ontology files at <http://geneontology.org>. A set of PHP scripts updates these GO term pages once per week as described in the **Supplementary Methods** section. MediaWiki automatically generates links to the child terms as subcategories, and links to associated genes as category members. We modified the normal MediaWiki category-page display with a set of extensions that provide an AJAX-based expansion/contraction to allow viewing the descendants of each child term. Users also have the option to filter associated genes for terms that have a large number of members. Each page also includes a link to the term in AmiGO, the official web-based term browser of the GO consortium. GONUTS term pages also include a graphical representation of that ontology node from the EMBL-EBI (4). Clicking on the graphic takes the user to the EBI's QuickGO browser (4,5). The key component on each GO term page is the area for user-editable usage notes. MediaWiki pages allow for associated 'Talk' pages, which provide a place for discussion, commentary and questions for other users. In GONUTS, the 'Talk' pages for GO terms are seeded with links to the Sourceforge tracker for the GO consortium, which allows users to find prior-related discussions.

Users can search for GO terms or genes with GO annotations using GO id numbers or keywords in the normal MediaWiki search system. Although MediaWiki search does not currently allow wildcards, the search will find complete words in any order. Thus, terms related to

cAMP-dependent protein kinases can be found using 'cAMP protein kinase' or 'kinase protein cAMP' in any order. To address the shortcomings of the native search, the ability to search GONUTS using Google has been added.

## USER-EDITABLE NOTES FOR EVERY GO TERM

What sets GONUTS apart from other ontology browsers is the editable notes section on every term page. Browsing and searching for ontology terms or annotated genes is unrestricted, but registration is required for editing the notes. This is mainly to inhibit vandalism of the wiki; 'wiki spam' is a serious problem on many open wikis (6–9). We use a 'vampire model' for user registration on GONUTS: any registered user can create additional registered users without applying for permission from the central resource. These, in turn, can register their colleagues. This method of user creation removes the dependence on the involvement of administrators and allows the community to grow dynamically.

Clicking the edit link to the right of the 'Notes' heading takes a logged-in user to a MediaWiki text entry page. Although editing the notes is not 'What You See Is What You Get' (WYSIWYG), the standard MediaWiki markup is easy to learn and will be familiar to users of Wikipedia and many other wikis. Help pages about editing are available on GONUTS, Wikipedia and across the web. An important aspect of all wiki-based systems is that there is no requirement for users to learn any of the markup in order to contribute or edit content. The collaborative nature of wiki editing means that information can be added in plain text by one user and then formatted by another who is more familiar with the markup system. Moreover, even incomplete comments are potentially useful starting points for elaboration by other users.

In order to maintain ease of use on the GO term pages, we have made a few modifications that make editing different from editing at Wikipedia. For example, we have simplified handling of references, which are important to the credibility of any community-edited text. A link is provided to insert a citation marker that accepts a PubMed ID number and retrieves the relevant reference information from PubMed's E-utilities (10). These are then collected in a references section that includes links to wiki pages for each publication.

## GENE PAGES

When we first conceived GONUTS we wanted examples of high-quality annotations using specific GO terms, which community curators could use as role models for their own annotation. For this purpose, we included all of the GO annotations from a set of well-curated model organisms from the GO consortium's reference genomes project (11,12). Gene pages generated for the reference genomes include links to a details page in the submitting model organism database, and lists of all the GO annotations associated with that gene. The references field is converted into a citation, and a references section lists the references

**(A.)** CACAO competition  
 The scoreboard is undergoing some testing. This should not affect annotating. Please email Daniel with any questions or comments about the upgrade.

**Category:GO:0016021 ! integral to membrane**

**Id:** GO:0016021  
**name:** integral to membrane  
**namespace:** cellular\_component  
**def:** Penetrating at least one phospholipid bilayer of a membrane. May also refer to the state of being buried in the bilayer with no exposure outside the bilayer. When used to describe a protein, indicates that "either part of the peptide sequence is embedded in the membrane." [GOC:go\_curators]  
**subset:** gosubset\_prok  
**synonym:** "transmembrane" RELATED [GOC:ma]  
**xref:** Wikipedia:Transmembrane\_protein  
**is\_a:** GO:0031224 ! intrinsic to membrane

**Last version checked** date: 28/10/2011 16:35 **Last updated** date: 06/11/2009 17:04  
 saved-by: gwg saved-by: midori auto-generated-by: OBO-Edit 2.0 auto-generated-by: OBO-Edit 2.0

**Gene Ontology Home** The contents of this box are automatically generated. You can help by adding information to the "Notes".

**Usage Notes**

**C.** A diagram illustrating membrane protein types: Intrinsic to membrane, Integral to membrane, Extrinsic to membrane, and Anchored membrane. It shows a lipid bilayer with proteins embedded in various ways.

**From the Cellular Component Ontology Guidelines**

**References** See Help:References for how to manage references in GONUTS. [edit]

**Child Terms** This term has the following 10 child terms, out of 10 total.  
 • [+] GO:0005887 ! integral to plasma membrane (41)  
 • [+] GO:0009319 ! cytochrome c oxidase complex  
 • [+] GO:0016472 ! sodium ion-transporting two-sector ATPase complex  
 • [+] GO:0017090 ! mephrin A complex  
 • [+] GO:0031301 ! integral to organelle membrane (10)  
 • [+] GO:0031361 ! integral to thylakoid membrane  
 • [+] GO:0034702 ! ion channel complex (4)  
 • [+] GO:0045203 ! integral to cell outer membrane  
 • [+] GO:0046696 ! lipopolysaccharide receptor complex  
 • [+] GO:0046930 ! pore complex (3)

**D.** Pages in category "GO:0016021 ! integral to membrane"  
 The following 200 pages are in this category, out of 22,415 total.

**E.** Category:  
 g  
 • GO:gosubset\_prok  
 i  
 • GO:0031224 ! intrinsic to membrane

This page was last modified on 9 November 2009, at 03:32. Content is available under . Privacy policy About GONUTS Disclaimers

**Figure 1.** A typical GO term page in GONUTS. (A) Information about the GO term derived from the ontology files from the GO consortium, including id, definition, relationships, and parent terms. (B) A ontology graph from the EBI. (C) User-edited section for notes. In this example, a user has embedded an uploaded diagram explaining the term in better detail. (D) Child Terms (E) Genes annotated to this term. The page also contains all the typical elements of a MediaWiki page including a sidebar for site-navigation and tabs along the top for various actions.

with links to PubMed and the citing database, as appropriate. As with the GO term pages, a Notes section is provided for user commentary. These notes about how the annotation of this gene was done by the source

MOD informs the annotation of similar genes in other organisms.

Although we initially envisioned GONUTS as an aid to annotation at other sites, such as EcoliWiki, we realized

that there is demand for community annotation of gene function for many genomes that do not have well-established model organism databases (13). To meet this need we implemented an automated gene-page creator. Clicking on ‘Create new gene page’ in GONUTS takes the user to a form where the desired gene can be specified by any accession UniProt understands (UniProt id, EMBL id, PDB id, etc.) Submitting the form launches a script that creates a gene page for the protein of interest and preloads it with various types of information. At present we do not support user-created pages for genes encoding RNA products; this is planned for a later version. Figure 2 shows one of these user-created gene pages. The top section of the page is constructed using gene names, synonyms and accessions from UniProt (14,15). GO annotations from UniProt are automatically placed in a table in the lower section of the gene page. This table is editable using our TableEdit MediaWiki extension.

Community curation can involve providing information that supports or refutes existing annotations or entering new GO annotations. In the former case, contributing can be as simple as adding a NOT qualifier, a PubMed ID to a table row, or text to the comments field. When adding new annotations, we imagine that the user will have another copy of GONUTS, AmiGO, or their favorite ontology browser open in another web browser window to find appropriate GO terms. The user need only know the GO id (the unique seven digit number) when adding an annotation as the TableEdit extension will fill in the appropriate term name and ontology code. Pulldown menus allow the user to choose qualifiers and evidence codes, and a custom extension recognizes when with/from identifiers are needed. References added as PMID:number are automatically converted to links and endnotes, with links to a literature page where users can document how the data in the linked paper supports each relevant GO annotation.

Each gene page is automatically added to the appropriate Category for each annotated GO term. Each gene page is also categorized by the taxonomy of the source organism using NCBI taxonomy. This allows searching for genes based on a GO term\_id and a taxonomy node. The gene pages are categorized for all parent nodes in the taxonomy tree, so that searches can be done at levels above species. This should be useful for comparing, for example, zebrafish and mouse annotations. A set of checkboxes controls which gene products are displayed; users can select any subset based on the source organism.

## GONUTS IS\_EDITED WEB SERVICE

GONUTS contains thousands of pages with content seeded from other resources. Pages where community members have added or edited content are currently a small subset of these. GONUTS provides a REST web service to allow other sites to display whether or not a GO term or gene page includes contributions from the community. The service, which can be accessed at [http://gowiki.tamu.edu/rest/is\\_edited.php](http://gowiki.tamu.edu/rest/is_edited.php), takes input parameters via HTTP GET that specify either a GO term identifier or a query string specifying a set of wiki pages. The

service returns the number of times the page has been revised by a registered user since it was created and the number of times it was edited after an optional date parameter. Further documentation of the web service is at [http://gowiki.tamu.edu/wiki/index.php/Help:Web\\_Services](http://gowiki.tamu.edu/wiki/index.php/Help:Web_Services). This service allows other web resources to incorporate intelligent links to GONUTS that display whether or not preexisting community annotation is present at the linked page. For example, AmiGO uses the web service to indicate whether user comments are present for GO terms (16). By using appropriate SQL wildcards in the page query, a list of all user-edited pages relating to a particular organism can be returned.

## ANNOTATION JAMBOREES WITH GONUTS

One of the strengths of wikis is the ability to easily adapt them to uses beyond their original design. In the Summer of 2008, GONUTS began hosting a series of electronic annotation jamborees for the RefGenome project (12) (see [http://gowiki.tamu.edu/wiki/index.php/Electronic\\_Jamboree](http://gowiki.tamu.edu/wiki/index.php/Electronic_Jamboree)). The purpose of these Jamborees was to compare the annotation practices of different groups and to develop stronger guidelines for annotation best practices. Genes targeted for annotation from all species are added to the jamboree’s page (a Mediawiki category page.) A custom extension aggregates all of the annotations into a table, and a web service from the Mouse Genome Informatics Database at Jackson Labs generates an image to compare annotations on a graphical representation of the ontologies (Figure 3) (17). Using these tools the annotators can see what others have done and compare annotations across both the phylogenetic tree and the Gene Ontology graph.

## USAGE OF GONUTS

Use of the GONUTS website as measured by Google Analytics has grown dramatically since its launch in 2007. GONUTS had approximately 4000 unique visitors in September of 2011 compared to just 1200 in September 2010. Although this is well below the usage of EcoliWiki, which had over 28 000 unique visitors during the same period, GONUTS is much more heavily edited than EcoliWiki. Since the two wikis launched, non-staff users account for only about 1500 page revisions in EcoliWiki. In contrast, non-staff GONUTS users have contributed almost 9700 revisions. In both cases, bot edits are excluded. Figure 4 shows a histogram showing editing activity by GONUTS users.

## DISCUSSION

Manually curated literature-based annotations using controlled vocabularies such as the Gene Ontology are the gold standard for providing information about gene function. However, the need for human curation of papers creates a serious bottleneck in annotation. GONUTS was created to support spreading the annotation effort across a larger community of biologists by providing

Log in 19:22:11GMT



Page Discussion Read View source View history Go Search

**CACAO competition**  
The scoreboard is undergoing some testing. This should not affect annotating. Please email Daniel with any questions or comments about the upgrade.

## ECOLI:ihfB

<b>Species (Taxon ID)</b>	Escherichia coli (strain K12). (taxon:83333)
<b>Gene Name(s)</b>	ihfB (synonyms: <i>himD</i> , <i>hip</i> )
<b>Protein Name(s)</b>	• Integration host factor subunit beta • IHF-beta
<b>External Links</b>	
UniProt Identifier	IHF_ECOLI
UniProt Accessions	POA6Y1, P08756,
EMBL	X04864, U00096, AP009048,
PIR	G64830,
RefSeq	AP_001542.1, NP_415432.1,
PDB	1IHF, 1OUZ, 1OWF, 1OWG, 2HT0, 2IE,
Pfam	PF00216,

**Contents [hide]**  
1 Annotations  
2 Notes  
3 References

### Annotations

Showing 1 to 48 of 48 entries Filter Rows: Evidence: Any/All

Qualifier	GO ID	GO term name	Reference	Evidence Code	with/from	Aspect	Notes	Status
	GO:0006350	transcription	PMID:8156991 <sup>[1]</sup>	IDA: Inferred from Direct Assay		P		
	GO:0003677	DNA binding	GO_REF:0000002	IEA: Inferred from Electronic Annotation	InterPro:IPR000119	F		
	GO:0016563	transcription activator activity	PMID:7629095 <sup>[2]</sup>	IDA: Inferred from Direct Assay		F		
	GO:0003677	DNA binding	GO_REF:0000002	IEA: Inferred from Electronic Annotation	InterPro:IPR020816	F		
	GO:0032079	endodeoxyribonuclease activity	PMID:7499339 <sup>[7]</sup>	IDA: Inferred from Direct Assay		F	Fig 3 & 5	complete

edit table

### Notes

### References

See Help:References for how to manage references in GONUTS.

- 1. ↑ 1.0 1.1 1.2 1.3 1.4 Zulianello L et al. (1994) The HimA and HimD subunits of integration host factor can specifically bind to DNA as homodimers. *EMBO J* **13**: 1534-40 PubMed GONUTS page
- 2. ↑ 2.0 2.1 2.2 2.3 2.4 Zulianello L et al. (1995) Participation of the flank regions of the integration host factor protein in the specificity and stability of DNA binding. *J Biol Chem* **270**: 17902-7 PubMed GONUTS page
- 3. ↑ 3.0 3.1 3.2 3.3 3.4 Mengertitsky G et al. (1993) Genetic and biochemical analysis of the integration host factor of Escherichia coli. *J Mol Biol* **231**: 646-57 PubMed GONUTS page
- 4. ↑ Presutti DG & Hassan HM (1995) Binding of integration host factor (IHF) to the Escherichia coli *sodA* gene and its role in the regulation of a *sodA-lacZ* fusion gene. *Mol Gen Genet* **246**: 228-35 PubMed GONUTS page
- 5. ↑ 5.0 5.1 5.2 5.3 5.4 Granston AE & Nash HA (1993) Characterization of a set of integration host factor mutants deficient for DNA binding. *J Mol Biol* **234**: 45-59 PubMed GONUTS page
- 6. ↑ Azam TA & Ishihama A (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**: 33105-13 PubMed GONUTS page
- 7. ↑ Nelson WC et al. (1995) The *trαY* gene product and integration host factor stimulate Escherichia coli DNA helicase I-catalyzed nicking at the F plasmid oriT. *J Biol Chem* **270**: 28374-80 PubMed GONUTS page
- 8. ↑ Sut MV et al. (2009) Protein and DNA effectors control the TrαI conjugative helicase of plasmid R1. *J Bacteriol* **191**: 6888-99 PubMed GONUTS page

<b>Category:</b>			
<b>c</b>	<b>p</b>	<b>s</b>	
• GO:0005694 / chromosome	• GO:0045941 / positive regulation of transcription	• GO:0043585 / sequence-specific DNA binding	
<b>d</b>	<b>r</b>	<b>t</b>	
• GO:0003677 / DNA binding	• GO:0045449 / regulation of transcription	• GO:0006351 / transcription, DNA-dependent	
• GO:0032058 / DNA duplex unwinding	• GO:0006355 / regulation of transcription, DNA-dependent	• GO:0016563 / transcription activator activity	
• GO:0006310 / DNA recombination	• GO:0006417 / regulation of translation	• GO:0016564 / transcription repressor activity	
<b>n</b>		<b>u</b>	
• GO:0016481 / negative regulation of transcription		• GO:0006350 / transcription	

This page was last modified on 19 July 2011, at 14:19.  
Content is available under .  
Privacy policy About GONUTS Disclaimers

Downloaded from <http://nar.oxfordjournals.org/> at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

Figure 2. A user-created GONUTS page for *E. coli ihfB* generated from its UniProt record.

## A Category:SMAD signaling

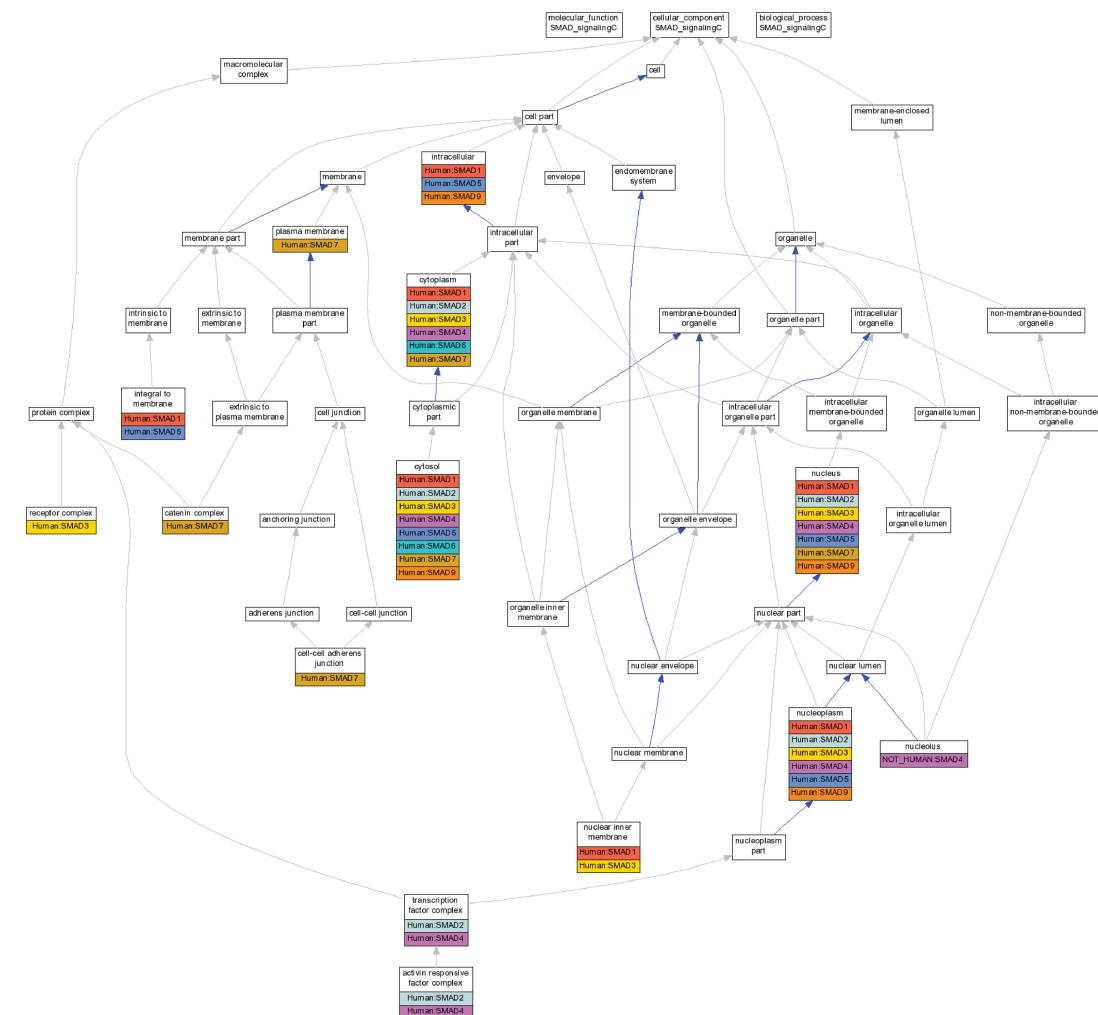
Annotation of human SMAD signaling proteins started in Dec 2008 as part of the cardiovascular GO annotation initiative, at UCL. For more information please contact v.khodiyar@ucl.ac.uk

Showing 1 to 125 of 125 entries

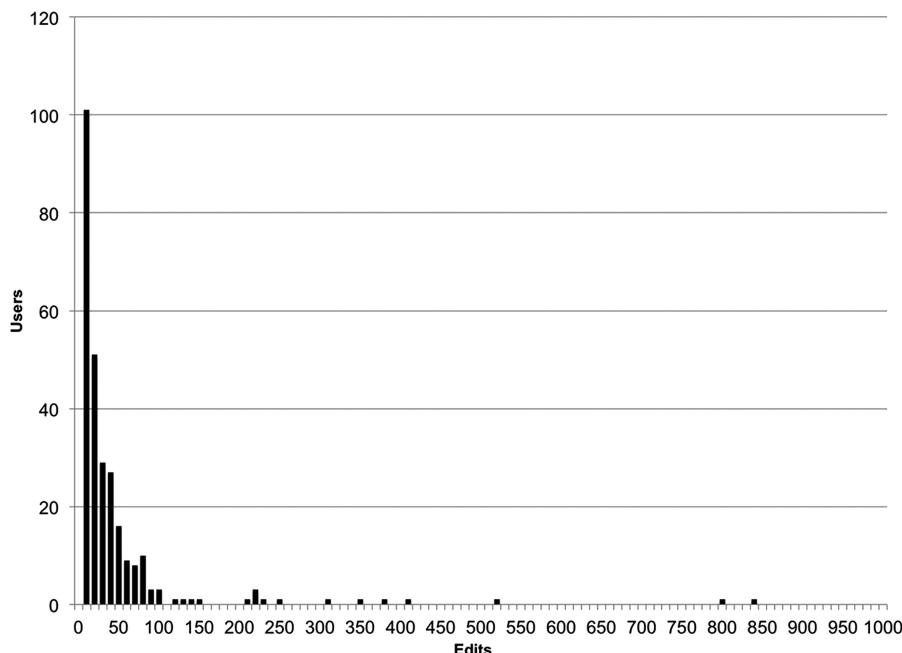
Search all columns:

Aspect	ID	Term	HUMAN SMAD1	HUMAN SMAD2	HUMAN SMAD3	HUMAN SMAD4	HUMAN SMAD5	HUMAN SMAD6	HUMAN SMAD7	HUMAN SMAD8
Biological Process	GO:0000122	negative regulation of transcription from R...			IDA IMP				IDA	
Biological Process	GO:0001666	response to hypoxia			IMP	IMP				
Biological Process	GO:0001707	mesoderm formation		ISS						
Biological Process	GO:0001933	negative regulation of protein phosphorylation			IMP					
Biological Process	GO:0006351	transcription, DNA-dependent	IDA	IDA	IDA	IDA				
Biological Process	GO:0006355	regulation of transcription, DNA-dependent	IDA							
Biological Process	GO:0006810	transport			IDA					
Biological Process	GO:0006917	induction of apoptosis			IMP					
Biological Process	GO:0006919	activation of caspase activity			IMP					
Biological Process	GO:0006955	immune response			IMP				IMP	
Biological Process	GO:0007050	cell cycle arrest			IMP					

## B



**Figure 3.** Screenshots of a user-created annotation jamboree category page for human SMAD signaling proteins started as part of the cardiovascular GO annotation initiative. Genes of interest are placed in the jamboree category. An extension generates a sortable and filterable table (A) and diagrams showing the annotations made to genes in this category (B). The table is truncated in this figure and only one of the three diagrams is shown. The full page can be viewed at ([http://gowiki.tamu.edu/wiki/index.php/Category:SMAD\\_signaling](http://gowiki.tamu.edu/wiki/index.php/Category:SMAD_signaling)).



**Figure 4.** Distribution of user contributions to GONUTS. Contributed revisions are grouped in bins of 10. Users with no edits are not shown.

both detailed documentation for GO terms and infrastructure for unofficial annotation and reannotation of any protein in UniProt (15). Although we ultimately envision GONUTS as a means for supporting broad community annotation, we have been pleased that GONUTS has turned out to be useful even for professional curators in the GO consortium who have access to a wide variety of alternative tools.

One of the basic features of wikis is the flexibility with which users can not only modify content on existing pages, but also create and organize pages in ways that were not anticipated by the site owners. We were thus especially pleased when we found that users had created categories to organize focused annotation efforts. The collaborative nature of wikis is another fundamental part of GONUTS, and over the past few years this feature has led several other groups to create wiki-based systems for biology (18–21). The free text nature of the basic wiki software limits the usefulness of user-created content. The use of custom tables, interfaces and markup tools in GONUTS gently guides users to create content that can be more easily mined than basic wikitext. Our approach could potentially be merged with existing semantic mediawiki tools used by others to integrate with the larger semantic web. Even without these, however, we can already easily create web services and data dumps for usage notes and annotations.

GONUTS also differs from most biology wikis in our extensive use of categories to reflect the DAG structure of the knowledge encoded in the Gene Ontology. The approach used in GONUTS could be easily adapted for other biological ontologies that share the obo file format and relationship types. Indeed, we have provided GONUTS loading code to the Sequence Ontology (SO)

project to help them generate their wiki-based term browser (22).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Methods.

## ACKNOWLEDGEMENTS

The authors thank Barry Wanner and Paul Thomas, as PIs of the EcoliHub and PortEco projects that provided the subcontracts to support this work. We thank Christine Elsik for introducing us to the Gene Ontology and the many members of the GO consortium for their support and feedback.

## FUNDING

EcoliWiki is funded as a component of PortEco from sub-contracts from grant U24 GM077905-01 (2006–2009) and 1U24GM088849-01 (2009–present) from the National Institutes of Health/National Institutes of General Medical Sciences. Funding for open access charge: National Institutes of Health (1U24GM088849).

*Conflict of interest statement.* None declared.

## REFERENCES

- Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

2. Gene Ontology Consortium. (2001) Creating the gene ontology resource: design and implementation. *Genome Res.*, **11**, 1425–1433.
3. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
4. Barrell,D., Dimmer,E., Huntley,R.P., Binns,D., O'Donovan,C. and Apweiler,R. (2009) The GOA database in 2009—an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.*, **37**, D396–D403.
5. Huntley,R.P., Binns,D., Dimmer,E., Barrell,D., O'Donovan,C. and Apweiler,R. (2009) QuickGO: a user tutorial for the web-based Gene Ontology browser. *Database (Oxford)*, **2009**, bap010.
6. Adler,B.T., de Alfaro,L., Mola-Velasco,S.M., Rosso,P. and West,A.G. (2011) Wikipedia vandalism detection: combining natural language, metadata, and reputation features. *Lect. Notes Comput. Sci.*, **6609**, 277–288.
7. Chin,S.-C., Street,W.N., Srinivasan,P. and Eichmann,D. (2010) Detecting Wikipedia vandalism with active learning and statistical language models. *Proceedings of the 4th Workshop on Information Credibility*. ACM, Raleigh, NC, USA, pp. 3–10.
8. Harpalani,M., Hart,M., Singh,S., Johnson,R. and Choi,Y. (2011) Language of vandalism: Improving Wikipedia vandalism detection via stylometric analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers – Volume 2*. Association for Computational Linguistics Portland, OR, USA, pp. 83–88.
9. West,A.G., Kannan,S. and Lee,I. (2010) Detecting Wikipedia vandalism via spatio-temporal analysis of revision metadata. *Proceedings of the Third European Workshop on System Security*. ACM, Paris, France, pp. 22–28.
10. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
11. Gene Ontology Consortium. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.
12. Reference Genome Group of the Gene Ontology Consortium. (2009) The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. *PLoS Comput. Biol.*, **5**, e1000431.
13. Bidartondo,M.I. (2008) Preserving accuracy in GenBank. *Science*, **319**, 1616.
14. UniProt Consortium. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
15. UniProt Consortium. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–D219.
16. Carbon,S., Ireland,A., Mungall,C.J., Shu,S., Marshall,B. and Lewis,S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.
17. Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E., Eppig,J.T. and the Mouse Genome Database Group (2011) The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.*, **39**, D842–D848.
18. Hodis,E., Prilusky,J. and Sussman,J.L. (2010) Proteopedia: a collaborative, virtual 3D web-resource for protein and biomolecule structure and function. *Biochem. Mol. Biol. Educ.*, **38**, 341–342.
19. Huss,J.W. III, Lindenbaum,P., Martone,M., Roberts,D., Pizarro,A., Valafar,F., Hogenesch,J.B. and Su,A.I. (2010) The Gene Wiki: community intelligence applied to human gene annotation. *Nucleic Acids Res.*, **38**, D633–D639.
20. Hoffmann,R. (2008) A wiki for the life sciences where authorship matters. *Nat. Genet.*, **40**, 1047–1051.
21. Florez,L.A., Roppel,S.F., Schmeisky,A.G., Lammers,C.R. and Stüke,J. (2009) A community-curated consensual annotation that is continuously updated: the *Bacillus subtilis* centred wiki SubtiWiki. *Database (Oxford)*, **2009**, bap012.
22. Eilbeck,K., Lewis,S.E., Mungall,C.J., Yandell,M., Stein,L., Durbin,R. and Ashburner,M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.